Question 2:

(a)(i)

assume 1 is incorrectly classified points and 0 is correctly classified points

```
[0, 0, 0, 0, 0, 0, 0, 1, 0, 0]
```

As we see, the probability of error doesn't equal 0.2 (it equals 0.1 in this case) and this is because there is a 0.8 probability that the model will correctly classify a point (i.e. outputs 0), so it is not certain that we get 8 zero's and 2 ones. It is even possible the model outputs all zeros since there is a high probability for a point to be correctly classified, which is 0.8.

(a)(ii)

for u=0.2 and N=10 :

The probability that the model outputs 2 errors out of 10 can be calculated using binomial distribution, that is:

$P(E_{D^{(10)}}(h) = \mu)$ = (10C2)*0.8^8 * 0.2^2=0.301989888

which is 0.30 after rounding.

(b)(i)

$\max\left\{E_{\mathcal{D}^{(10)}}(h)\right\},$	$\min\left\{E_{\mathcal{D}^{(10)}}(h)\right\},$	sample mean $\left\{E_{\mathcal{D}^{(10)}}(h)\right\},$	sample standard deviation $\left\{E_{\mathcal{D}^{(10)}}(h)\right\}.$
0.5	0	0.206	0.11984990613262905

(ii)

there are 68 runs with probability of error different than 0.2, which means we have 32 runs with probability of error of 0.2 . It quite agrees with the theoretically calculated probability, which was 0.3. In this case, we got 0.32 (32 runs out of 100) which is very close but not equal, we need more runs in order for the simulated probability to converge to the theoretical one.

(iii) estimated $P(|E_{D^{(10)}}(h) - \mu| < 0.05)$ = 0.32

c(i)

for u=0.2 and N=100 :

The probability that the model outputs 20 errors out of 100 can be calculated using binomial distribution, that is:

$P(E_{D^{(10)}}(h) = \mu)$ = (100C20)*0.8^80 * 0.2^20=0.0993002148088246896 38

which is 0.10 after rounding.

(b)(i)

$\max\{E_{\mathcal{D}^{(10)}}(h)\}$,	$\min\{E_{\mathcal{D}^{(10)}}(h)\}$,	sample mean $\{E_{\mathcal{D}^{(10)}}(h)\}$,	sample standard deviation $\{E_{\mathcal{D}^{(10)}}(h)\}$.
0.36	0.12	0.203	0.04517742799230607

(ii)

there are 88 runs with probability of error different than 0.2, which means we have 12 runs with probability of error of 0.2 . It quite agrees with the theoretically calculated probability, which was 0.1. In this case, we got 0.12 (12 runs out of 100) which is very close but not equal, we need more runs in order for the simulated probability to converge to the theoretical one.

(iii) estimated $P(|E_{D^{(10)}}(h) - \mu| < 0.05)$ = 0.7

c(ii)

for u=0.5 and N=10:

The probability that the model outputs 5 errors out of 10 can be calculated using binomial distribution, that is:

$P(E_{D^{(10)}}(h) = \mu)$ = (10C5)*0.5^5 * 0.5^5=0.24609375

which is 0.25 after rounding.

(b)(i)

$\max\left\{E_{\mathcal{D}^{(10)}}(h)\right\}$,	$\min\left\{E_{\mathcal{D}^{(10)}}(h)\right\}$,	sample mean $\left\{E_{\mathcal{D}^{(10)}}(h)\right\}$,	sample standard deviation $\left\{E_{\mathcal{D}^{(10)}}(h)\right\}$.
0.8	0.1	0.4960000000000001	0.154220621189256

(ii)

there are 76 runs with probability of error different than 0.5, which means we have 24 runs with probability of error of 0.5 . It quite agrees with the theoretically calculated probability, which was 0.25. In this case, we got 0.24 (24 runs out of 100) which is very close but not equal, we need more runs in order for the simulated probability to converge to the theoretical one.

(iii) estimated $P\left(\left|E_{D^{(10)}}(h) - \mu\right| < 0.05\right)$ = 0.24

c(iii)

for u=0.5 and N=100:

The probability that the model outputs 50 errors out of 100 can be calculated using binomial distribution, that is:

$P(E_{D^{(10)}}(h) = \mu)$ = (100C50)*0.5^50 * 0.5^50=0.0795892373871787

which is 0.08 after rounding.

(b)(i)

$\max\left\{E_{\mathcal{D}^{(10)}}(h)\right\},$	$\min\left\{E_{\mathcal{D}^{(10)}}(h)\right\},$	sample mean $\left\{E_{\mathcal{D}^{(10)}}(h)\right\},$	sample standard deviation $\left\{E_{\mathcal{D}^{(10)}}(h)\right\}.$
0.61	0.39	0.49670000000000003	0.049780618718533426

(ii)

there are 92 runs with probability of error different than 0.5, which means we have 8 runs with probability of error of 0.5 . It agrees with the theoretically calculated probability, which was 0.08. In this case, we got 0.08 (8 runs out of 100) which is equal to the theoretically calculated probability.

(iii) estimated $P\left(\left|E_{D^{(10)}}(h) - \mu\right| < 0.05\right)$ = 0.64

(d)(i)

The estimate of the error rate gets more accurate as we increase N. For example, the sample mean error went from 0.206 to 0.203 when we increased N from 10 to 100.

(ii) For N=10, there were 33 datasets where $E_{D(10)}(h) \leq 0.45$, and there were only 19 datasets for N=100. The vast majority of the datasets, However, have probability of error more than 0.45, which indicated the classifier didn't learn anything